

Metric spaces: The definition, and some examples

Stephen Semmes
Rice University

A *metric space* is a set M together with a real-valued function $d(x, y)$ defined for $x, y \in M$ that satisfies the following three conditions. First, $d(x, y) \geq 0$ for every $x, y \in M$, and $d(x, y) = 0$ if and only if $x = y$. Second,

$$(1) \quad d(x, y) = d(y, x)$$

for every $x, y \in M$. Third,

$$(2) \quad d(x, z) \leq d(x, y) + d(y, z)$$

for every $x, y, z \in M$, which is known as the *triangle inequality*. This function $d(x, y)$ is called the *metric* on M , and represents a measurement of distance between elements of M .

Remember that the *absolute value* $|x|$ of a real number x is equal to x when $x \geq 0$, and is equal to $-x$ when $x \leq 0$. As usual,

$$(3) \quad |xy| = |x| |y|$$

and

$$(4) \quad |x + y| \leq |x| + |y|$$

for every $x, y \in \mathbf{R}$. The standard metric on the real line is defined by

$$(5) \quad d(x, y) = |x - y|,$$

which is easily seen to satisfy the requirements of metric mentioned before.

These informal notes have been prepared in connection with a lecture at the high school mathematics tournament held at Rice University on February 20, 2010.

The standard Euclidean metric on \mathbf{R}^n is defined by

$$(6) \quad d_2(x, y) = \left(\sum_{j=1}^n (x_j - y_j)^2 \right)^{1/2},$$

where $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$. This clearly satisfies the positivity and symmetry conditions for a metric, but the triangle inequality is less obvious. The latter is a well-known theorem in classical Euclidean geometry, and can also be shown using the Cauchy–Schwarz inequality. Alternatively, it is easy to check that

$$(7) \quad d_1(x, y) = \sum_{j=1}^n |x_j - y_j|$$

and

$$(8) \quad d_\infty(x, y) = \max_{1 \leq j \leq n} |x_j - y_j|$$

also define metrics on \mathbf{R}^n , using the triangle inequality for the standard metric on the real line. Although these three metrics on \mathbf{R}^n are different geometrically, they are equivalent topologically, which basically means that they are the same in terms of continuity and convergence.

If $(M, d(x, y))$ is any metric space, then the (open) ball $B(x, r)$ with center $x \in M$ and radius $r > 0$ is defined by

$$(9) \quad B(x, r) = \{x \in M : d(x, y) < r\}.$$

In the case of \mathbf{R}^n equipped with the standard Euclidean metric, $B(x, r)$ is an ordinary round ball. However, if \mathbf{R}^n is equipped with the metric $d_1(x, y)$, then $B(x, r)$ is diamond-shaped. If \mathbf{R}^n is equipped with the metric $d_\infty(x, y)$, then $B(x, r)$ is a cube with sides parallel to the coordinate axes.

If $(M, d(x, y))$ is any metric space and $0 < \alpha \leq 1$, then one can show that $d(x, y)^\alpha$ also defines a metric on M . The main point is to check that the triangle inequality still holds, by verifying that

$$(10) \quad (a + b)^\alpha \leq a^\alpha + b^\alpha$$

for every pair a, b of nonnegative real numbers. Equivalently,

$$(11) \quad a + b \leq (a^\alpha + b^\alpha)^{1/\alpha}$$

for every $a, b \geq 0$. If $1/\alpha$ is an integer, then this follows by expanding the right side into a sum. Otherwise, (10) can be seen as a nice exercise in calculus. One can also use

$$(12) \quad \max(a, b) \leq (a^\alpha + b^\alpha)^{1/\alpha}$$

to get that

$$(13) \quad a + b \leq (a^\alpha + b^\alpha) \max(a, b)^{1-\alpha} \leq (a^\alpha + b^\alpha)^{1/\alpha}.$$

As before, $d(x, y)^\alpha$ is equivalent to $d(x, y)$ topologically, even if they may be different geometrically.

If M is any set, then one can define a metric on M by putting $d(x, y) = 1$ when $x \neq y$, and equal to 0 when $x = y$. It is easy to see that this satisfies the requirements of a metric, which is known as the *discrete metric* on M .

Let p be a prime number, and let $x = (a/b)p^j$ be a rational number, where a, b, j are integers, $a, b \neq 0$, and a, b are not divisible by p . The *p-adic absolute value* of x is denoted $|x|_p$ and defined to be p^{-j} under these conditions, and to be 0 when $x = 0$. One can check that

$$(14) \quad |xy|_p = |x|_p |y|_p$$

and

$$(15) \quad |x + y|_p \leq \max(|x|_p, |y|_p) \leq |x|_p + |y|_p$$

for every pair of rational numbers x, y . The *p-adic metric* on the set \mathbf{Q} of rational numbers is given by

$$(16) \quad d_p(x, y) = |x - y|_p.$$

In particular,

$$(17) \quad d_p(x, z) \leq \max(d_p(x, y), d_p(y, z)) \leq d_p(x, y) + d_p(y, z)$$

for every $x, y, z \in \mathbf{Q}$. This is quite different from the standard metric on \mathbf{Q} , since $|x|_p \leq 1$ for every integer x , and $|x|_p$ may be quite small even when x is a nonzero integer. The *p-adic numbers* are obtained by completing the rational numbers with respect to the *p-adic metric*, in the same way that the real numbers can be obtained by completing the rational numbers with respect to the standard metric.

If $(M, d(x, y))$ is any metric space and E is a subset of M , then the restriction of $d(x, y)$ to $x, y \in E$ is a metric on E . In this way, every subset of a metric space may be considered as a metric space too.

Let \mathbf{S}^n be the unit sphere in \mathbf{R}^{n+1} with respect to the standard Euclidean metric, so that

$$(18) \quad \mathbf{S}^n = \left\{ (x_1, \dots, x_{n+1}) \in \mathbf{R}^{n+1} : \sum_{j=1}^{n+1} x_j^2 = 1 \right\}.$$

If $x, y \in \mathbf{S}^n$, $x \neq y$, and $x \neq -y$, then there is a unique 2-dimensional plane $P(x, y)$ in \mathbf{R}^{n+1} that passes through x , y , and 0 . Thus

$$(19) \quad C(x, y) = P(x, y) \cap \mathbf{S}^n$$

is a circle with radius 1, and we let $d_{\mathbf{S}^n}(x, y)$ be the length of the shorter arc on $C(x, y)$ that connects x to y . If $x = y$, then we simply put $d_{\mathbf{S}^n}(x, y) = 0$. If $x = -y$, then x, y lie on the same line through 0 , and we can use any plane $P(x, y)$ that contains this line. In this case, all of the circular arcs on \mathbf{S}^n that connect x to y are half-circles, and we take $d_{\mathbf{S}^n}(x, y)$ to be their common length, which is π . It is well known that $d_{\mathbf{S}^n}(x, y)$ satisfies the triangle inequality, and this will be discussed further later on. Using this, it is easy to see that $d_{\mathbf{S}^n}(x, y)$ defines a metric on \mathbf{S}^n , and that this metric is topologically equivalent to the restriction of the standard Euclidean metric on \mathbf{R}^{n+1} to \mathbf{S}^n .

Suppose now that M is some kind of nice surface in \mathbf{R}^n , which may be of any dimension. Suppose also that M is connected in the sense that for every pair of points x, y in M there is a continuously-differentiable path on M that goes from x to y . In this situation, it is natural to try to define the distance between x and y to be the length of the shortest curve on M that goes from x to y . In particular, it can be shown that such a curve exists under suitable conditions. Alternatively, one can avoid the issue by defining the distance from x to y to be the infimum or greatest lower bound of the lengths of the paths on M that go from x to y . It is easy to see that this automatically satisfies the triangle inequality, using the fact that a path from x to y may be combined with a path from y to z to get a path from x to z whose length is equal to the sum of the lengths of the other two paths for any $x, y, z \in M$. If M is the unit sphere, then it is well known that this is the same as the metric described in the previous paragraph.

Consider the space $C([0, 1])$ of continuous real-valued functions on the unit interval $[0, 1]$. One way to define a metric on $C([0, 1])$ is by

$$(20) \quad d_1(f, g) = \int_0^1 |f(x) - g(x)| dx.$$

Another way is to use

$$(21) \quad d_\infty(f, g) = \max_{0 \leq x \leq 1} |f(x) - g(x)|.$$

Note that the maximum of $|f(x) - g(x)|$ is attained on $[0, 1]$, by the extreme value theorem. It is not difficult to check that $d_1(f, g)$ and $d_\infty(f, g)$ both satisfy the requirements of a metric on $C([0, 1])$, and that

$$(22) \quad d_1(f, g) \leq d_\infty(f, g)$$

for every $f, g \in C([0, 1])$.

The unit sphere

Let us return to the distance function $d_{\mathbf{S}^n}(x, y)$ defined on the unit sphere \mathbf{S}^n in \mathbf{R}^{n+1} as before. Note that $d_{\mathbf{S}^n}(x, y)$ attains its maximal value π when x and y are antipodal, which means that $x = -y$. Let $\|x\| = \left(\sum_{j=1}^{n+1} x_j^2\right)^{1/2}$ be the standard Euclidean norm of $x = (x_1, \dots, x_{n+1}) \in \mathbf{R}^{n+1}$, so that $\|x - y\|$ is the standard Euclidean metric on \mathbf{R}^{n+1} . The spherical distance on \mathbf{S}^n is related to the standard Euclidean metric on \mathbf{R}^{n+1} by

$$(23) \quad \sin\left(\frac{d_{\mathbf{S}^n}(x, y)}{2}\right) = \frac{\|x - y\|}{2},$$

which holds for every $x, y \in \mathbf{S}^n$. This can be verified by considering the line through 0 and the mid-point $(x + y)/2$ of the line segment that connects x and y , which are perpendicular to each other.

If $x, y, z \in \mathbf{S}^n$, then we would like to show that

$$(24) \quad d_{\mathbf{S}^n}(x, z) \leq d_{\mathbf{S}^n}(x, y) + d_{\mathbf{S}^n}(y, z).$$

We may as well suppose that $x \neq y \neq z$ and $d_{\mathbf{S}^n}(x, y) + d_{\mathbf{S}^n}(y, z) < \pi$, since this is trivial otherwise. In particular, $d_{\mathbf{S}^n}(x, y), d_{\mathbf{S}^n}(y, z) < \pi$, which implies that $x, z \neq -y$. Let $P(x, y)$ be the 2-dimensional plane in \mathbf{R}^{n+1} passing

through x , y , and 0 , and let $C(x, y)$ be the circle which is the intersection of $P(x, y)$ with \mathbf{S}^n , as before. If z is an element of $C(x, y)$, then (24) is clear.

Consider

$$(25) \quad \begin{aligned} \Sigma &= \{w \in \mathbf{S}^n : \|y - w\| = \|y - z\|\} \\ &= \{w \in \mathbf{S}^n : d_{\mathbf{S}^n}(y, w) = d_{\mathbf{S}^n}(y, z)\}, \end{aligned}$$

which is a sphere of dimension $n - 1$ in \mathbf{R}^{n+1} . The intersection of Σ with $C(x, y)$ consists of exactly two points u , v , and we can label them in such a way that

$$(26) \quad \|x - u\| \leq \|x - v\|,$$

which is equivalent to

$$(27) \quad d_{\mathbf{S}^n}(x, u) \leq d_{\mathbf{S}^n}(x, v).$$

The main point now is that

$$(28) \quad \|x - u\| \leq \|x - w\| \leq \|x - v\|$$

for every $w \in \Sigma$, which implies that

$$(29) \quad d_{\mathbf{S}^n}(x, u) \leq d_{\mathbf{S}^n}(x, w) \leq d_{\mathbf{S}^n}(x, v).$$

In particular, we can apply this to $w = z$, to get that

$$(30) \quad d_{\mathbf{S}^n}(x, z) \leq d_{\mathbf{S}^n}(x, v).$$

This implies (24), since we already know that (24) holds when $z \in C(x, y)$. To get (28), note that Σ is centered at a point σ on the line segment connecting y to $-y$, and that Σ is contained in the n -dimensional plane H in \mathbf{R}^{n+1} that passes through σ and is perpendicular to the line $L(y)$ through y and 0 . Let L' be the line contained in $P(x, y)$ that passes through σ and is perpendicular to $L(y)$, which is the same as the intersection of $P(x, y)$ with H . Also let x' be the orthogonal projection of x in L' , so that $x' \in L'$ and the line L'' passing through x and x' is perpendicular to L' . If x is already an element of L' , then $x = x'$. Note that L'' is parallel to $L(y)$, and that x' is the same as the orthogonal projection of x in H . Thus

$$(31) \quad \|x - w\|^2 = \|x - x'\|^2 + \|x' - w\|^2$$

for every $w \in H$, so that maximizing or minimizing $\|x - w\|$ for $w \in \Sigma$ is the same as maximizing or minimizing $\|x' - w\|$ on Σ . The maximum and

minimum of $\|x' - w\|$ for $w \in \Sigma$ are attained on the line L' , since L' passes through x' and the center σ of Σ and is contained in H . We also have that

$$(32) \quad L' \cap \Sigma = P(x, y) \cap \Sigma = C(x, y) \cap \Sigma,$$

where the first step uses the fact that $L' = P(x, y) \cap H$, the second step uses the fact that $C(x, y) = P(x, y) \cap \mathbf{S}^n$, and both steps use the fact that $\Sigma = \mathbf{S}^n \cap H$. It follows that the maximum and minimum of $\|x' - w\|$ on Σ are attained on $C(x, y) \cap \Sigma$, which are the same as the maximum and minimum of $\|x - w\|$ on Σ , as desired.

Infinite series

If x is a real number, then it is well known and easy to see that

$$(33) \quad (1 - x) \sum_{j=0}^n x^j = 1 - x^{n+1}$$

for each nonnegative integer n . Here x^n is interpreted as being equal to 1 when $n = 0$, even when $x = 0$. In particular,

$$(34) \quad \sum_{j=0}^n x^j = \frac{1 - x^{n+1}}{1 - x}$$

when $x \neq 1$. If $|x| < 1$, then $|x^n| = |x|^n \rightarrow 0$ as $n \rightarrow \infty$, and so

$$(35) \quad \lim_{n \rightarrow \infty} \sum_{j=0}^n x^j = \frac{1}{1 - x}.$$

Equivalently, $\sum_{j=0}^{\infty} x^j$ converges as an infinite series of real numbers, and

$$(36) \quad \sum_{j=0}^{\infty} x^j = \frac{1}{1 - x}.$$

Suppose now that p is a prime number, x is a rational number, and $|x|_p < 1$. This implies that $|x^n|_p = |x|_p^n \rightarrow 0$ as $n \rightarrow \infty$, which means that $x^n \rightarrow 0$ as $n \rightarrow \infty$ with respect to the p -adic metric on \mathbf{Q} . Similarly,

$$(37) \quad \lim_{n \rightarrow \infty} \sum_{j=0}^n x^j = \frac{1}{1 - x}$$

with respect to the p -adic metric on \mathbf{Q} . This is the same as saying that $\sum_{j=0}^{\infty} x^j$ converges as an infinite series of rational numbers with respect to the p -adic metric, with the same sum $1/(1-x)$ as before. For example, one can apply this to $x = p$, to get that $\sum_{j=0}^{\infty} p^j$ converges with respect to the p -adic metric, with sum equal to $1/(1-p)$.

If an infinite series $\sum_{j=1}^{\infty} a_j$ of real numbers converges, then it is well known that $a_j \rightarrow 0$ as $j \rightarrow \infty$ with respect to the standard metric on \mathbf{R} . It is also well known that $\sum_{j=1}^{\infty} a_j$ may not converge even though $\lim_{j \rightarrow \infty} a_j = 0$, e.g., when $a_j = 1/j$. By contrast, it can be shown that an infinite series $\sum_{j=1}^{\infty} a_j$ of p -adic numbers converges if and only if $a_j \rightarrow 0$ as $j \rightarrow \infty$ with respect to the p -adic metric. In both cases, convergence of an infinite series is defined to mean convergence of the corresponding sequence of partial sums, which is equivalent to asking that the sequence of partial sums be a Cauchy sequence, by completeness. In the p -adic case, it is much easier to check that a sequence is a Cauchy sequence, because of the stronger form of the triangle inequality.

Norms on \mathbf{R}^n

A real-valued function $\|x\|$ on \mathbf{R}^n is said to be a *norm* if it satisfies the following three conditions. First, $\|x\| \geq 0$ for every $x \in \mathbf{R}^n$, with $\|x\| = 0$ if and only if $x = 0$. Second,

$$(38) \quad \|tx\| = |t| \|x\|$$

for every $t \in \mathbf{R}$ and $x \in \mathbf{R}^n$, where $tx = (tx_1, \dots, tx_n)$. Third,

$$(39) \quad \|x + y\| \leq \|x\| + \|y\|$$

for every $x, y \in \mathbf{R}^n$, where $x + y = (x_1 + y_1, \dots, x_n + y_n)$. If $\|x\|$ is a norm on \mathbf{R}^n , then it is easy to see that

$$(40) \quad d(x, y) = \|x - y\|$$

defines a metric on \mathbf{R}^n .

For example,

$$(41) \quad \|x\|_2 = \left(\sum_{j=1}^n x_j^2 \right)^{1/2}$$

is the standard Euclidean norm on \mathbf{R}^n , for which the corresponding metric is the standard Euclidean metric $d_2(x, y)$, defined earlier. Similarly, it is easy

to see that

$$(42) \quad \|x\|_1 = \sum_{j=1}^n |x_j|$$

and

$$(43) \quad \|x\|_\infty = \max_{1 \leq j \leq n} |x_j|$$

are also norms on \mathbf{R}^n , and that these norms correspond to the metrics $d_1(x, y)$ and $d_\infty(x, y)$ that were mentioned at the beginning.

A subset E of \mathbf{R}^n is said to be *convex* if for every $x, y \in E$ and $t \in \mathbf{R}$ with $0 \leq t \leq 1$, we have that

$$(44) \quad tx + (1 - t)y \in E.$$

This is the same as saying that the line segment in \mathbf{R}^n connecting x to y is contained in E for every $x, y \in E$. Let $\|x\|$ be a norm on \mathbf{R}^n , and let

$$(45) \quad B = \{x \in \mathbf{R}^n : \|x\| \leq 1\}$$

be the closed unit ball in \mathbf{R}^n corresponding to $\|x\|$. It is easy to see that B is convex. Moreover, B is also *symmetric* about the origin in \mathbf{R}^n , in the sense that $-x \in B$ for every $x \in B$.

The definition of a norm makes sense on any vector space V over the real numbers, such as the space $C([0, 1])$ of continuous real-valued functions on the unit interval. Any norm on V determines a metric on V in the same way as before, and the closed unit ball associated to the norm is convex and symmetric about the origin in V .

Distances on graphs

A *graph* G can be described by a set V of *vertices*, and a subset E of the set of subsets of V with exactly two elements. Thus we say that there is an *edge* between two distinct elements v, w in V when $\{v, w\} \in E$. A finite sequence v_0, v_1, \dots, v_n of vertices in G is said to determine a *path* in G if there is an edge between v_j and v_{j+1} for each $j = 0, \dots, n-1$, in which case the length of this path is defined to be n . We say that G is *connected* if for every pair of vertices $v, w \in V$ there is a path v_0, \dots, v_n with $v_0 = v$ and $v_n = w$. If $v = w$, then we can simply take $n = 0$ and $v_0 = v = w$.

If G is a connected graph and $v, w \in V$, then the distance $d(v, w)$ between v and w in G may be defined to be the smallest nonnegative integer n for

which there is a path in G of length n that goes from v to w . It is easy to see that this defines a metric on V . In particular, if $v, w, z \in V$, then any path in G from v to w can be combined with a path from w to z to get a path from v to z , whose length is the sum of the lengths of the paths from v to w and from w to z . This implies that this definition of distance satisfies the triangle inequality.

This definition can be extended so that the distance between adjacent vertices is any positive real number, depending on the vertices. The edges can also be represented by segments or other curves, and included in the associated metric space. In this case, one can consider continuous paths in the metric space, and not just discrete paths.

Metric spaces are often discussed in books on basic analysis and topology, in connection with the theory behind calculus. A few references along these lines are given below.

References

- [1] R. Beals, *Analysis: An Introduction*, Cambridge University Press, 2004.
- [2] R. Goldberg, *Methods of Real Analysis*, 2nd edition, Wiley, 1976.
- [3] F. Gouvêa, *p -Adic Numbers: An Introduction*, 2nd edition, Springer-Verlag, 1997.
- [4] B. Mendelson, *Introduction to Topology*, 3rd edition, Dover, 1990.
- [5] F. Morgan, *Real Analysis*, American Mathematical Society, 2005.
- [6] F. Morgan, *Real Analysis and Applications*, American Mathematical Society, 2005.
- [7] S. Krantz, *The Elements of Advanced Mathematics*, 2nd edition, Chapman & Hall / CRC, 2002.
- [8] S. Krantz, *Real Analysis and Foundations*, 2nd edition, Chapman & Hall / CRC, 2005.
- [9] M. Rosenlicht, *Introduction to Analysis*, Dover, 1986.
- [10] W. Rudin, *Principles of Mathematical Analysis*, 3rd edition, McGraw-Hill, 1976.